

The research program of the Center for Economic Studies (CES) produces a wide range of theoretical and empirical economic analyses that serve to improve the statistical programs of the U.S. Bureau of the Census. Many of these analyses take the form of CES research papers. The papers are intended to make the results of CES research available to economists and other interested parties in order to encourage discussion and obtain suggestions for revision before publication. The papers are unofficial and have not undergone the review accorded official Census Bureau publications. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the U.S. Bureau of the Census. Republication in whole or part must be cleared with the author(s).

**NEW EVIDENCE ON SEX SEGREGATION AND SEX DIFFERENCES IN WAGES  
FROM MATCHED EMPLOYEE-EMPLOYER DATA\***

Kimberly Bayard  
University of Maryland  
Center for Economic Studies, U.S. Census Bureau

Judith Hellerstein  
University of Maryland  
NBER

David Neumark  
Michigan State University  
NBER

Kenneth Troske  
University of Missouri

CES 98-18      December 1998

All papers are screened to ensure that they do not disclose confidential information. To obtain a copy, submit comments, or obtain general information about the series, contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Washington Plaza II, Room 211, Bureau of the Census, Washington, DC 20233-6101, (301-457-1882) or INTERNET address

snguyen@ces.census.gov.

### Abstract

We assemble a new matched employer-employee data set covering essentially all industries and occupations across all regions of the U.S. We use this data set to re-examine the question of the relative contributions to the overall sex gap in wages of sex segregation vs. wage differences by sex within occupation, industry, establishment, and occupation-establishment cells. This new data set is especially useful because earlier research on this topic relied on data sets that covered only a narrow range of industries, occupations, or regions. Our results indicate that a sizable fraction of the sex gap in wages is accounted for by the segregation of women into lower-paying occupations, industries, establishments, and occupations within establishments. Nonetheless, a substantial part of the sex gap in wages remains attributable to the individual's sex. This latter finding contrasts sharply with the conclusions of previous research (especially Groshen, 1991), which indicated that sex segregation accounted for essentially all of the sex wage gap. Further research into the sources of within-establishment within-occupation sex wage differences is therefore much more important than previously thought.

\*Bayard is a Ph.D. student in economics at the University of Maryland and a Social Science Analyst at the Center for Economic Studies, U.S. Census Bureau. Hellerstein is Assistant Professor of Economics at the University of Maryland, and a Faculty Research Fellow of the NBER. Neumark is Professor of Economics at Michigan State University, and a Research Associate of the NBER. Troske is Assistant Professor of Economics at the University of Missouri. We thank Fran Blau, Bill Evans, Erica Groshen, Robert Lalonde, and Brooks Pierce for helpful comments. This research was supported by NSF grant SBR95-10876 through the NBER. This paper was originally prepared for the U.S. Bureau of the Census's International Symposium on Linked Employer/Employee Data, Washington, D.C., 1998. The research in this paper was conducted while the last three authors were research associates with the Center for Economic Studies, U.S. Bureau of the Census. Research results

and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Bureau of the Census or the Center for Economic Studies.

## I. Introduction

Women have consistently earned lower wages than men in U.S. labor markets, although this gap has narrowed in recent decades (Blau, 1998). Understanding the sources of sex differences in wages is vital to determining why the wage gap between men and women persists. Previous research has focused on the effect of the occupational segregation of men and women on the wage gap (e.g. Macpherson and Hirsch, 1995), the effect of industry segregation (e.g. Fields and Wolff, 1995), and to a lesser extent on the segregation of men and women into different employers (Blau, 1977; Bielby and Baron, 1984; Carrington and Troske, 1998).<sup>1</sup> These studies all find evidence that the wage gap falls considerably once one accounts for segregation.

Evidence on the contribution to the wage gap of within-establishment within-occupation segregation is far harder to find. Indeed, we are not aware of any empirical work on this issue that uses large data sets representative of a wide array of industries. The reason for this is the paucity of data sets containing detailed demographic information for multiple workers in the same establishment. As a result, studies of the effects of establishment and occupation-establishment segregation have used unusual, somewhat narrow data sets. The most complete study of which we are aware is Groshen (1991). She uses various Industry Wage Surveys (IWS) conducted between 1974 and 1983 to estimate the proportions of the sex difference in wages attributable to the percent female in the occupation, the establishment, and the job-cell (i.e., within occupations within establishments), in five specific industries (Miscellaneous Plastic Products, Nonelectrical Machinery, Life Insurance, Banking,

---

<sup>1</sup>Blau also documents the role of segregation by industry.

and Computer and Data Processing).<sup>2</sup> She finds that the segregation of women into lower-paying occupations within employers also contributes to the wage gap. Groshen finds little remaining wage gap between men and women once she accounts for the relationship between wages and the percent female within establishments, occupations, and establishment-occupation cells.<sup>3</sup> Groshen concludes, then, that pay differences between men and women cannot be a result of wage discrimination by employers of the type that violates the Equal Pay Act.

Groshen's results suggest that understanding the persistence of the sex wage gap amounts to understanding the forces that lead to sex segregation in the workforce. Such forces could range from employer discrimination in hiring and promotion to different preferences across the sexes for the non-pecuniary benefits of different jobs. Groshen is well aware of the possible limited representativeness of her data, however, noting that, "although theoretically ideal, study of a full national cross-section of industries is impossible because of the lack of appropriate data" (p. 460).

In this paper we construct and use an extended version of the Worker-Establishment Characteristics Database (WECD) to decompose the source of male/female wage differentials.<sup>4</sup> Like the WECD, this data set matches workers who responded to the long form of the 1990 Decennial Census to the establishments where they work, using the U.S. Census Bureau's Standard Statistical

---

<sup>2</sup>The IWS originally covered 70 industries, but Groshen reports results only for these five. These surveys have been discontinued. Her work builds on Blau's (1977), but provides a more complete decomposition of the sex gap in wages, including evidence on the importance of an individual's sex within occupation, establishment, and job cell. Blau studies BLS Area Wage Surveys conducted in 1970, which covered clerical, professional, and technical occupations; she focuses on a subset of these occupations in three large northeastern cities.

<sup>3</sup>This mirrors to some extent Blau's (1977) earlier finding that "intraoccupational pay differentials by sex are primarily the result of differences in wage-rates among firms rather than within firms" (p. 73).

<sup>4</sup>For a description of the original WECD, see Troske (forthcoming).

Establishment List (SSEL). However, whereas the WECD is limited to manufacturing plants, this new data set (the New Worker-Establishment Characteristics Database, or NWECD) includes workers and establishments from all sectors of the economy. Thus, we provide new estimates of the role of various dimensions of sex segregation in generating sex differences in wages, based on a large data set with workers in nearly all regions, industries, and occupations.<sup>5</sup>

Our results indicate that a sizable fraction of the sex gap in wages is accounted for by the segregation of women into lower-paying occupations, industries, establishments, and occupations within establishments. We also find, however, that a very substantial part of the sex gap in wages remains attributable to the individual's sex. These findings contrast sharply with the conclusions of previous research (especially Groshen, 1991) indicating that sex segregation accounted for essentially all of the sex wage gap. We explore a number of possible reasons for the differences in results. We ultimately conclude that the data we use yield different answers from the data Groshen used, *even for the same occupations and industries that she studied*.

## II. The Data

The data used in this study come from a match between worker records from the 1990 Sample Detail File (SDF) to establishment records in the 1990 Standard Statistical Establishment List (SSEL).<sup>6</sup> The 1990 SDF consists of all household responses to the 1990 Decennial Census long form. As part of the Decennial Census, five-sixths of all households receive a "short-form," which asks seven basic demographic questions about household members ("person questions") and seven questions about the

---

<sup>5</sup>Because of the matching described below, and other sample restrictions, the data set is not nationally representative.

<sup>6</sup>Both the SDF and the SSEL are confidential Census Bureau products, as is the NWECD that we describe in this paper.

housing unit ("housing questions"). The remaining one-sixth receive a more extensive "long-form" survey, which contains 26 additional person and 19 additional housing questions. These questionnaires ask households to supply distinct information for each household member. Those receiving the long-form are asked to identify each employed household member's (1) occupation, (2) employer location, and (3) employer's industry in the previous week. The Census Bureau then assigns occupational, industrial, and geographic codes to long-form responses. Thus, the SDF contains the standard demographic information for workers collected on the long form of the Decennial Census, along with detailed location information and a three-digit census industry code for each respondent's place of work.

The SSEL is an annual list of business establishments maintained by the U.S. Census Bureau. The SSEL contains detailed location information and a four-digit SIC code for each establishment, along with a unique establishment identifier that is common to other Census Bureau economic surveys and censuses. It also includes information on total payroll expenses, employment, and whether or not the establishment is part of a multi-establishment firm.

We matched workers and establishments using the detailed location and industry information available in both data sets. Briefly, the first step in the matching process was to keep only establishments unique to an industry-location cell.<sup>7</sup> Next, all workers indicating that they work in the same industry-location cell as a retained establishment were linked to the establishment. The matched data set is the NWECD. Because the SDF contains only a sample of workers, and because not all workers are matched, the matched data set includes a sample of workers at each establishment.

---

<sup>7</sup>We do this because we do not actually have the employer name available on both establishment and worker records.

The Census Bureau organizes the United States into different geographic areas, assigning codes to each. For the NWECD, there are five areas of interest: state, county, place, tract, and block.<sup>8</sup> The geographic coding process works primarily as a hierarchy. The Census Bureau assigns unique codes to every state in the country. Within states, each county is also assigned a unique code. In addition, in areas or townships with a population of 2,500 or more, the Census Bureau assigns a place code. Because an area or town can cross county boundaries, we can distinguish between areas in the same place but different counties. Tract codes are unique within counties and block codes are unique within tracts. The Census Bureau uses the same geographic codes in both the SSEL and the Decennial Census.

Two shortcomings of the geographic codes in the SSEL are (1) the absence of tract and block codes before 1992, and (2) the incomplete assignment of these codes to all establishments. To assign tract and block codes to the 1990 SSEL, we extracted each establishment's block and tract code (when available) from the 1992 SSEL, and then matched these codes back to the 1990 SSEL. Some establishments that ceased operation between 1990 and 1992 do not appear in the 1992 SSEL, making it impossible to identify block and tract codes for these establishments. Not all establishments had tract and block codes assigned as of 1992 due to address problems. In the 1992 SSEL, Census had assigned tract and block codes to 45% of all establishments.

In addition to geographic codes, the Census Bureau assigns industry codes to the SDF and the SSEL. The Census Bureau asks long-form respondents to identify their employer's industry, which the Census Bureau codes into one of 244 Census Industry Classification (CIC) codes. Each CIC

---

<sup>8</sup>In some geographic areas, the Census Bureau uses Block Numbering Areas (BNAs) instead of tracts. For our purposes, a BNA is equivalent to a tract. The Census Bureau assigns tracts and blocks in tandem, so whenever an establishment is assigned a tract code, it is also always assigned a block code.



corresponds roughly to a three-digit Standard Industrial Classification (SIC) code.<sup>9</sup> In the SSEL, the Census Bureau assigns each establishment a six-digit SIC code based on the plant's primary economic activity.<sup>10</sup> Since the CIC codes are more aggregated than SIC codes, we use a concordance table to assign a CIC to each SIC in the SSEL.<sup>11</sup>

The first step in the matching process is to assign industry-location cells to all plants in the SSEL. We divide the SSEL into plants that are unique in a state-county-place-industry (SCPI) cell and those that are not, and retain all unique SCPI plants. In cases where there are multiple plants in an SCPI cell, we first retain the cell only if all plants in the cell have tract and block codes. We then keep only those plants that are unique within a state-county-place-tract-block-industry cell.

Next, we assign workers in the SDF to industry-location cells based on information provided in the SDF. Unlike the SSEL, the Census Bureau assigns detailed geography codes to all observations in the SDF.<sup>12</sup> Once we have workers assigned to industry-location cells and have establishments that are unique within a cell, we can match the workers to the particular establishments where they work.

We take a number of additional steps to improve the quality of the match. First, to ensure that workers are matched properly to employers in the WECD, we discard all workers and establishments

---

<sup>9</sup>An exception is Construction. There is one CIC for Construction and this corresponds to the equivalent of three two-digit SIC codes.

<sup>10</sup>The last two digits of the SIC code are product codes for goods-producing industries, or type of business codes for service establishments.

<sup>11</sup>A few SICs correspond to more than one CIC. We omitted establishments in these industries.

<sup>12</sup>When long-form respondents omit geographic information, the Census Bureau imputes missing values.

from the matched sample where Census imputed either the worker's or the establishment's industry.<sup>13</sup>

We also discard all workers from the matched sample if the worker's place-of-work code is imputed and if this imputed code is the source of the match.<sup>14</sup>

Second, some matches lead to apparent inconsistencies, prompting us to discard matches when the number of workers matched to an establishment exceeds the number of employed workers as reported by the establishment in the SSEL. There are several reasons why the number of matched workers might exceed total employment. First, there may be errors in the industry or geographic codes for some workers or establishments in the SDF or SSEL. Second, there is a time lag between when the Census Bureau surveys workers and employers. Census asks workers where they worked on April 1 and employers how many workers they employed as of March 12; total employment on April 1 may exceed total employment on March 12. A third problem is that workers may be incorrectly assigned to locations because of imprecise SDF questions. Because the SDF asks workers only where they worked in the past week, workers who were working at a site other than their employer's primary location may be improperly assigned to an establishment. Fourth, in the SSEL total employment includes only a plant's employees, not its owners. In the SDF, however, both owners and employees are assigned to a particular establishment. Thus, although there may be legitimate reasons for the number of matched workers to exceed reported establishment employment, to avoid potentially

---

<sup>13</sup>This imputation occurs for plants when an incomplete SIC code (only the first two or three digits) is provided. For such cases, the Census Bureau randomly assigns the remaining digits.

<sup>14</sup>To understand the exclusion based on imputed geographic data, consider the following example. When a worker is matched to an establishment unique in an SCPI cell, the match is based on the state, county, place, and industry of the worker and the establishment. If the worker's block code is imputed, then this imputed code has no bearing on the match, and we retain the match in the data. However, if the match relies on tract and block-level information, and the worker's place-of-work block code is imputed, then we discard the worker from the matched data set.

incorrect matches we discard cases where this occurs.

In our matched sample, we also imposed some restrictions on both individuals and establishments. We include only individuals who report usually working between 30 and 65 hours per week, and 30 or more weeks in the last year (1989). These restrictions on hours and weeks are intended to pick out full-time, full-year workers who are less likely to have changed jobs in the past year, as well as those whose hours are so high that they may have held multiple jobs. We make these restrictions for three reasons. First, because the Decennial Census collects data on earnings from all jobs, rather than wages on the current job, we need to try to eliminate variation in wages that stems from multiple job-holding at a point in time or during the previous calendar year.<sup>15</sup> Second, because the 1990 Decennial Census asks workers to report the address of the establishment where they worked in the previous week, while the earnings data are for the previous calendar year, job changing may lead to inaccurate measurement of earnings in the matched data. Finally, Groshen's IWS data, with which we are trying to draw some comparisons, cover only full-time workers. We also restrict the sample to workers aged 18 to 65, with a constructed hourly wage ((annual earnings/weeks worked)/usual hours worked per week) in the range \$2.50 to \$500, and those working in establishments in public administration (in order to restrict our focus to the private sector).

We also require that establishments have total employment of at least 25, for two reasons: first, when we compared average establishment-level worker earnings in the matched observations in the SDF with average payroll expenses in the SSEL, these corresponded much more closely for establishments with 25 or more workers; second, the IWS industry samples included mainly

---

<sup>15</sup>Multiple job holding rates are virtually identical among men and women. 1996 CPS data indicate rates of 6.2% for men and 6.1% for women (Stinson, 1997). Thus, although multiple job holding could affect our data, it is unlikely to influence the sex differences we estimate.

establishments with 25 or more workers, and we are interested in a comparison of our results to Groshen's results using the IWS. In addition, to ensure that we have a reasonable basis for estimating the characteristics of an establishment's workforce, we required that the number of matched workers be least five percent of employment as reported in the SSEL. Finally, we eliminated the less than .1% of establishments that reported earnings exceeding more than \$600,000 per worker.

Table 1 documents the effects of these various matching rules and exclusion restrictions on the sample size, the number of matched workers, and average earnings and employment calculated from both the SSEL and SDF data. We define measures of establishment earnings per worker from data in both the SSEL and the SDF. For the SSEL, earnings per worker is constructed as Total Annual Payroll/Total Employment. For the SDF, establishment earnings per worker is created by averaging the annual wages and salaries of all workers matched to the establishment. The table shows that seven percent of establishments can be assigned to unique industry-location cells. These establishments are nearly twice as large as those in the overall sample (with an average of 41.17 workers, vs. 21.10 for the full SSEL sample), but have average earnings that are lower by about \$2200. This does not contrast with standard size-wage effects (Brown and Medoff, 1989), since there are no controls for industry, etc., and the ability to assign establishments to unique industry-location cells is not random with respect to these characteristics.

The next three rows (c-e) provide information on the observations on workers in the SDF. Out of a total of 17,311,211 workers, we match 1.1 million, or 6.5%, to establishments, once we discard unreliable matches or workers without earnings data. There are, of course, numerous establishments to which no workers are matched, reflected in the decline in the number of matched establishments from 388,787 to 201,944 based on the simple match, and 156,332 once other restrictions are imposed.

Naturally, the establishments to which workers in the SDF are matched tend to be larger, with average employment of 83.24. The last three columns compare earnings data. Average establishment earnings per worker reported in the SSEL are about \$1200 lower than the corresponding figure estimated from the SDF (\$18,218 vs. \$19,416); this is presumably attributable to the fact that in the SDF individuals can report earnings from more than one job. We also find, comparing columns (6) and (7) of row e, that average earnings per worker in the SDF data are about \$3,170 higher than average establishment earnings estimated from the same (matched) data. These numbers can differ because the earnings per establishment figures are not weighted by the number of matched workers in estimating average establishment earnings per worker; thus, this result likely stems from the concentration of higher-earning workers in larger establishments. Row f drops workers based on the restrictions on hours, weeks, age, wages, etc., with little impact except to drop those with lower earnings.

The final sample is described in row g, which we obtain after dropping establishments with fewer than 25 employees, with an insufficiently small percentage of matched employees, and with earnings outliers. We end up with a sample of 637,718 workers matched to 32,931 establishments. These establishments are of course much larger than those represented in the previous rows, and have an average of 16.14 workers matched to them. We also find that among this subset of larger establishments, average establishment earnings estimated from the SSEL and the SDF are considerably closer (\$20,983 vs. \$23,328).

Descriptive statistics for the matched sample are reported in column (1) of Table 2. The sample is approximately 47% female and seven percent black, with an average age of 40. The percentage ever married is 86. Because in estimating wage equations we want to account for the possible effects of children on women's wages, in particular, we extracted from the SDF the number of

children ever born, which is only available for women, and therefore defined to be zero for men.<sup>16</sup> On average, women in the NWECD have had 1.84 children. With respect to education, 21.1% have a bachelor's degree or higher, and 50 % report no college education. Column (3) reports descriptive statistics for the entire SDF file, with the weeks and hours restrictions imposed. Most of the demographic characteristics are quite close in the matched and full sample, with the exception of the cumulative fertility rate (for women) in the matched sample. Geographically, individuals living in MSA's are less likely to be in the matched sample, presumably because in urban areas individuals are less likely to work in unique industry-location cells. Turning to occupation, laborers are over-represented in the matched sample, and support occupations under-represented. Similarly, the industry composition of the sample is heavily weighted toward manufacturing, with 52% of workers in this industry, vs. 24% in the full sample, while retail is grossly under-represented, presumably because most retail establishments are in locations in which similar establishments are located (such as malls). In the empirical analysis, we address the potential consequences of the over-representation of manufacturing establishments.

The fact that the NWECD does not comprise a representative sample of U.S. workers is not surprising given the requirements for a match, and given the size restrictions imposed on matched establishments. For our purposes, however, the important question is whether the NWECD is not representative in ways that will bias the wage regressions we estimate. To partially answer this question, we report in Table 3 estimates from basic wage regressions with and without industry and occupation controls. Columns (1) and (2) provide benchmark estimates from wage regressions with

---

<sup>16</sup>In the regression analysis that follows, this implies that we simply restrict the effect of children ever born on men's wages to be zero.

and without industry and occupation controls, respectively, using workers from the SDF.<sup>17</sup> Not surprisingly, the results from the SDF are very similar to those from other large, nationally representative data sets (such as the CPS). The male-female wage gap in column (1) is 23.1%, but falls to 17.4% in column (2) when we control for broad occupation and industry categories. Similarly, the black-white wage gap is significant in both regressions, but it is smaller in column (2). Both columns show evidence of quadratic age profiles and positive returns to education, although the returns to education are smaller in column (2). Columns (3) and (4) of Table 3 replicate the specifications of columns (1) and (2), respectively, but use the NWECD data. The male-female wage gap in column (3) is 28.2%, which is slightly larger than that in the SDF (23.1%), but the difference in the male-female wage gap between the two data sets is virtually eliminated once we control broadly for industry and occupation (17.4% in the SDF vs. 16.6% in the NWECD). There are some other minor differences looking across the columns of Table 3 between the two data sets, but for the most part the wage regression results from the NWECD come close to replicating those from the SDF, particularly once controls for industry and occupation are added. Note that in the empirical work below, we always include some sort of controls for industry and occupation since one of our interests is in the effects of industry and occupation segregation on male-female wage differences.

Therefore, while the NWECD data are not representative of the underlying population of U.S. workers, the NWECD represents a clear improvement over existing data sources used to study the role of sex segregation along a number of dimensions in the workplace. The NWECD covers essentially the entire array of industries, occupations, locations, etc., in the U.S. economy. Moreover,

---

<sup>17</sup>Actually, because of computing constraints and the very large sample size of the SDF (1.7 million workers), the data used to construct columns (1) and (2) of Table 3 come from a random ten percent sample of the SDF.

wage regression estimates from the NWECD do not differ substantively from those obtained from a representative sample of the U.S. population of workers.

### III. Methods

In our initial empirical work, we assume that the wage gap between men and women is a function of individual human capital characteristics and characteristics of the "femaleness" of where a worker works, as represented by the percent female in a worker's occupation, industry, establishment, and occupation within an establishment (job cell). That is, we estimate wage regressions of the following form:

$$(1) \quad \ln(w_{poiej}) = \alpha + \beta F_p + (\gamma_1 OCC\%F_o + \gamma_2 IND\%F_i + \gamma_3 EST\%F_e + \gamma_4 JOB\%F_j + X_{poiej}M + \epsilon_{ijkmn} ,$$

where  $w$  is the log hourly wage,  $F$  is a dummy variable equal to one if individual  $p$  is female,  $OCC\%F$  is the percent female in occupation  $o$ ,  $IND\%F$  the percent female in industry  $i$ ,  $EST\%F$  the percent female in establishment  $e$ , and  $JOB\%F$  the percent female in job cell  $j$ . A vector of control variables is represented by  $X$ .

With the estimated coefficients of equation (1) in hand, we can construct standard wage decompositions that decompose the difference in average log wages between women and men (denoted  $w'_f$  and  $w'_m$ ) as follows:

$$(2) \quad w'_f - w'_m = \alpha' + (\gamma_1' (OCC\%F_f - OCC\%F_m) + \gamma_2' (IND\%F_f - IND\%F_m) + \gamma_3' (EST\%F_f - EST\%F_m) + \gamma_4' (JOB\%F_f - JOB\%F_m) + (X_f - X_m)M' ,$$

where primes on the coefficients indicate estimates, and 'f' and 'm' subscripts on the variables indicate means for women and men, respectively. This decomposition gives us the proportion of the wage gap due to individual sex differences in wages for workers in the same job cell ( $\alpha'$ ), and due to segregation of women into lower-wage occupations ( $(\gamma_1' [OCC\%F_f - OCC\%F_m])$ ), industries ( $(\gamma_2' [IND\%F_f -$



IND%F<sub>m</sub>]), establishments (8'[EST%F<sub>f</sub> - EST%F<sub>m</sub>]), and job cells (2'[JOB%F<sub>f</sub> - JOB%F<sub>m</sub>]), and differences in other observable characteristics. These decompositions can therefore be thought of as traditional Oaxaca (1973) decompositions, imposing the restriction that the coefficients are the same for men and women. We present most of our results imposing this restriction,<sup>18</sup> but also report key results below using the unrestricted decomposition.

While establishments are well-defined, industry and occupation can be defined at a variety of levels of disaggregation. Since the question of primary concern is within- vs. across-job wage differences, we are interested in trying to narrow occupational classifications. If, however, we use highly disaggregated occupations, we can end up with very small job cells (establishment-occupation cells), particularly since we only have a sample of workers in each plant. This may cause measurement error problems, which we discuss below. Consequently, we report evidence from specifications using a variety of levels of occupational disaggregation, beginning with 13 broad Census occupations, and then using successive levels of disaggregation of occupations used by the Census Bureau, down to the finest level of disaggregation into 505 occupations (of which 491 are represented in our data). The Census coding of occupations corresponds to some extent to 3-digit Standard Occupation Classification (SOC) codes.<sup>19</sup> To preview the results, we find that while using different levels of occupational disaggregation does change the quantitative results, the qualitative results are not strongly affected by the level of occupational detail that we use.

The percent-female variables in equation (1) are all estimated directly from the data. The

---

<sup>18</sup>This specification follows Groshen (1991).

<sup>19</sup>We combined a small number of Census occupations. Because we do not look at establishment-industry cells (since all workers in an establishment are presumably in the same industry, and are so by construction in our data set), we face no constraint in disaggregating industries finely, and hence we always use the detailed Census industry codes.

percentages female in the occupation and industry are estimated from the full SDF sample, so measurement error is likely to be minimal. However, the percentages female in the plant and job cell are estimated by necessity from the matched data in the NWECD. On average 16.14 workers are matched to a plant, so job-cell estimates, in particular, are often based on a small number of observations. Measurement error in these estimates therefore could be sizable, biasing the estimates of  $\delta$  and  $\gamma$  towards zero (and presumably biasing the other coefficient estimates as well, although a priori the direction of bias is unclear). Since our research focuses on the relative importance of sex differences in wages within vs. across occupation, industry, establishment, and job cell (i.e., the relative magnitudes of  $\delta$  vs.  $\gamma$ ,  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$ ), this measurement error is of substantive importance. To assess the influence of measurement error on the estimates, we first estimate the regressions and decompositions for the full sample, and then for restricted samples of individuals who work in job cells with a minimum of 10 observations per cell.

We also report results in which we estimate  $\delta$  controlling for fixed occupation, industry, establishment, and job-cell effects, rather than controlling for the percent female in each of these categories; this amounts, of course, to putting in job-cell dummy variables since these absorb occupation, industry, and establishment effects. In the absence of measurement error, we would not expect estimates of  $\delta$  obtained using these fixed effects to differ much from estimates using the percent-female variables if the percent-female variables do a reasonable job of characterizing how wages are affected by the sorting of workers into different industries, occupations, establishments, and job cells. Using job-cell dummies, however, avoids the measurement error inherent in the percent-female variables, and therefore should provide more reliable estimates of the within-job-cell sex difference in wages ( $\delta$ ). Nonetheless, most of the results we report use the percent-female variables, both to follow some of the

earlier literature (e.g., Groshen, 1991), and because the estimated effects of the percent-female variables are of interest in their own right—for example, in inferring the potential effects of a policy of comparable worth (Johnson and Solon, 1986).

The next section presents the empirical analysis of the NWECD data. We first present our basic estimates, and then numerous analyses that address measurement error bias and other estimation issues. Our estimates turn out to differ sharply from Groshen's (1991) earlier estimates of the roles of sex segregation and within-job sex differences in wages, so in the following section we explore in detail the differences between these alternative estimates.

#### IV. Results Using the NWECD

##### *Basic Analysis*

Table 4 begins by reporting results of wage regression estimations using 13 broad occupation categories.<sup>20</sup> To estimate the raw wage gap, column (1) simply reports estimates from a regression of log hourly wages on the female dummy variable; the raw gap in these data is -.375.<sup>21</sup> Column (2) then reports wage regression estimates introducing the four percent-female variables, but no other controls. Controlling for segregation by industry, occupation, establishment, and job cell, the sex gap in wages falls by about one-third, to -.244. Wages are lower in establishments with a higher percent female, and

---

<sup>20</sup>These 13 occupation categories are: (1) managerial and professional specialty--executive, administrative and managerial ; (2) managerial and professional specialty--professional specialty; (3) technical sales and administrative support--technicians and related support; (4) technical sales and administrative support--sales; (5) technical and administrative support--administrative support; (6) service--private household; (7) service--protective; (8) service--except private household and protective; (9) farming, forestry and fishing; (10) precision production, craft, and repair; (11) operators, fabricators, and laborers--machine operators; (12) operators, fabricators, and laborers--transportation and material moving; (13) operators, fabricators, and laborers--handlers, equipment cleaners, helpers, and laborers.

<sup>21</sup>All of our coefficient estimates are highly significant, so while standard errors are reported, we do not continually refer to statistical significance.

within establishments, in occupations with a higher percent female (the job-cell effect).<sup>22</sup> In this specification without other controls (notably education), occupational and industry segregation have the opposite of the usual effects, with wages higher in occupations and industries with a higher percent female. As discussed below, this result is generally reversed when individual-level controls are added. At the same time, it should be noted that most studies of sex segregation do not control for establishment and job cell (e.g., Macpherson and Hirsch, 1995; Johnson and Solon, 1986, Fields and Wolff, 1995); thus, these studies no doubt overstate the role of occupational and/or industry segregation per se. The decomposition also requires the calculation of mean differences between women and men in the right-hand-side variables, which are reported in column (3). As we would expect, women are in occupations, industries, establishments, and job cells with a higher percent female.

Columns (4) and (5) present the first decomposition, for the wage regression specification in column (2). Column (4) reports the absolute contribution of each variable to the wage gap, and column (5) the relative contribution. The estimates in column (5) indicate that over half (65.1%) of the wage gap is attributable to sex differences in wages that remain after accounting for segregation by occupation, industry, establishment, and job cell. Just over one-third (34.7%) is due to segregation into lower-paying occupations within establishments.

---

<sup>22</sup>Macpherson and Hirsch (1995) caution that the percent female in the worker's occupation may be a proxy for other job-related characteristics, so that the estimated effect on wages partially reflects compensating differentials based on workers' preferences, and perhaps also different skill requirements. Their evidence is consistent with this, as the longitudinal estimate of the effect of percent female in the occupation is much weaker than the cross-sectional estimate. Sorenson (1989) presents similar evidence for women only, based on a comparison of OLS estimates with estimates that account for selectivity into employment and into female-dominated occupations (although one can raise questions regarding identification of this model). In both papers, despite the evidence of bias, occupational segregation continues to lower wages.

Columns (6) and (7) report wage regression estimates and the "relative contribution" decomposition once we add human capital controls. Note that the Decennial Census does not have any direct information on experience or tenure; instead, we include age and its square. Because for women past childbearing is negatively associated with wages, especially in the absence of experience or tenure controls (Korenman and Neumark, 1992), we include a variable measuring the number of children ever born to women, which is set to zero for men. Given that this may also affect the rate of growth of wages over the life cycle, we also include an interaction of this variable with age (Polachek, 1975). The estimated effect of each additional child born is a two percent reduction in women's wages, while each child is associated with .5 percentage point lower wage growth over ten-year intervals. Thus, these variables display the qualitative effects of bearing and rearing children that we would expect.

Upon including these controls, along with those for education, marital status, race, and location, the estimated coefficient of the female dummy variable declines in absolute value (from -.244 to -.132), while the estimated coefficient of the percent female in the job cell declines by more (from -.243 to -.097). In this specification, the estimated effects of occupation and industry segregation become negative. In terms of the decomposition, after accounting for the effects of sex segregation by occupation, industry, establishment, and job cell, the sex difference in wages remains large, contributing over one-third of the sex gap (.351). The contribution of establishment segregation remains about the same, while the contribution of segregation within jobs within establishments falls by over half (to .138).

#### *The Effects of the Degree of Occupational Disaggregation*

The results in Table 4 are based on 13 highly-aggregated occupations. Because sex segregation (reflected in the mean difference in the proportion female in women's vs. men's

occupations) is likely to be more severe at a more-detailed occupational level, the decomposition results may be sensitive to the level of occupational aggregation used. To explore this question, in Table 5 we report results for increasing degrees of occupational disaggregation. Column (1) replicates the key results (for the specification including controls) from Table 4. In the second column, we increase the number of occupational classifications to 72, which amounts to disaggregating each of the 13 original occupations into anywhere from two to 14 distinct occupations. In the last column, we disaggregate as much as our data allow and use 491 of the most-detailed Census occupation codes. Each detailed Census occupation code corresponds generally to a mix of three-digit and four-digit Standard Occupation Classification (SOC) codes, often combining two or three four-digit occupations into a Census occupation.<sup>23</sup> The first two rows of the table show, as we would expect, that given the greater degree of sex segregation in more detailed occupations, the mean sex difference in the percent female by occupation and job cell is larger in each successive column. (The figures for industry and establishment are unchanged, of course.)

Turning to the wage regression estimates, the estimated coefficient of the female dummy variable, or the individual sex effect, declines as more detailed occupations are used, from  $-.132$  in column (1) to  $-.091$  in column (3), with the corresponding relative contribution to the wage gap falling from  $.351$  to  $.243$ . Nonetheless, a sizable sex wage gap persists. Of all the segregation measures, the contribution of job-cell segregation is most affected by the level of occupational

---

<sup>23</sup>For an example of what this occupational disaggregation really means, consider one of our 13 Census occupation codes, Technicians and Related Support Occupations. At the level of 72 total occupations, this category constitutes three separate occupations: (1) Health Technologists and Technicians; (2) Technologists and Technicians, Except Health Engineering and Related Technologists and Technicians; Science Technicians; and (3) Technicians, Except Health, Engineering, and Science. At the level of 491 total occupations, these three categories are further disaggregated into 22 distinct occupations, including such occupations as dental hygienists, survey and mapping technicians, and legal assistants.

disaggregation. The proportion of the sex gap accounted for by job-cell segregation rises from 13.8 to 23.6%.

### *Including Establishment-Level Controls*

To assess the robustness of our basic results and those using different levels of occupational disaggregation, we also estimated these same specifications including controls for establishment size and whether the establishment belongs to a multi-unit firm. To the extent that these reflect establishment-level characteristics, they may "over-control" for establishment-level differences, because they may capture dimensions of sex segregation. On the other hand, these variables may be related to unobserved human capital, calling for their inclusion along with the other human capital controls. Table 6 reports the results. They are very similar to those in Table 5, except that, as we might expect from controlling for some establishment characteristics that are correlated within an industry, the contribution of sex segregation by industry and establishment to the overall sex gap in wages falls slightly.

### *Manufacturing vs. Non-Manufacturing*

Earlier, we noted that the sample of establishments in the NWECD is heavily weighted towards manufacturing. If the effects of sex segregation are different in manufacturing and non-manufacturing industries, then in order to obtain unbiased estimates of population parameters, one might want to weight the data to make them more representative. Of course, many other factors influence selection into the NWECD sample, some of which may be related to unobservable characteristics, making it unclear how the weights might be constructed. In this subsection, therefore, we address the most salient non-representativeness--the preponderance of manufacturing plants--by reporting separate estimates for the manufacturing and non-manufacturing sectors.

Table 7 reports results of the relative contribution of sex itself and of each dimension of segregation to the sex wage gap, exactly paralleling the estimates reported in Table 5, although now disaggregated by sector. (The coefficient estimates are not shown.) There are some differences

between the two sectors. In general, there is more segregation in the non-manufacturing sector except at the job-cell level; this may reflect the greater variety of industries comprising this sector. In addition, there are some differences in the relative contributions to the wage gap of occupation, industry, and job-cell segregation. However, the key finding that the individual's sex accounts for a large share of the wage gap persists in both the manufacturing and non-manufacturing sectors. For example, in column (3), using the most disaggregated occupations, sex alone accounts for 24.8% of sex wage gap in manufacturing, and 23.3% in non-manufacturing. Thus, although this evidence does not address all sources of non-representativeness in the NWECD, it indicates that the most pronounced source of non-representativeness has little influence on the results.

### *Cohort Effects*

Although our human capital measures include controls for how many children a woman has had, and an interaction between children and age, these are only proxies for the labor market interruptions of women. To the extent that they do not fully capture the differential labor market experience of similarly aged women and men, our regression results, especially the coefficient on the female dummy, may be biased. To the extent that younger cohorts of women have had fewer labor market interruptions than older cohorts, this bias should be less severe for younger workers. Table 8 therefore replicates the specifications in Table 5, but uses only the sample of workers who are below the median age.<sup>24</sup> The raw wage gap between young men and women is smaller than that for the full sample (-.26 versus -.38), and the coefficient estimates on the female dummy variable are smaller (ranging from -.05 to -.08 versus -.09 to -.13). Because the extent of segregation between men and women is very similar for young workers relative to the full sample, however, and because the regression coefficients on the

---

<sup>24</sup>The segregation measures for percent female in the industry, occupation, establishment, and job cell are calculated from the full sample.



segregation variables are also very similar, the relative contribution of the female dummy to the wage gap for younger workers is not very different than for the full sample. Individual sex still accounts for anywhere between 20 and 31% of the wage gap. The results from the full sample are therefore not being driven by unmeasured differential labor market experiences of men and women.

### *The Effects of Measurement Error on the Estimates*

As discussed above, sampling error in the estimates of the percent female in the establishment and job cell may be quite severe. We do a number of things to assess the role of measurement error that this may induce. First, in Table 9 we report results from reestimating the specifications in Table 5 using data only on job cells with at least 10 workers. By restricting attention to larger job cells, we should reduce measurement error in the percent female in the job cell. Of course, this restriction also entails a cost in terms of the number of observations, particularly when more-detailed occupations are used. As the sample sizes towards the bottom of Table 9 show, this restriction results in excluding about 30% of the sample when we use the most aggregated occupations, rising to about 68% when we use the most-detailed occupational classification. Our estimates may also change because the restriction we impose tends to exclude smaller establishments.<sup>25</sup>

The bottom rows of Table 9 summarize the evidence on the contributions of each component of segregation and the individual's sex to the overall sex wage gap with and without the job-cell-size restriction. We generally obtain higher estimated effects of job-cell segregation, with the estimates nearly doubling and job-cell segregation contributing 46% of the sex wage gap for the most-detailed occupational breakdown. The estimated effects of establishment segregation change somewhat, while

---

<sup>25</sup>Although the observed characteristics of workers in this restricted sample are not vastly different from those in the full sample, workers in the restricted sample are slightly higher-paid and slightly more likely to be married.

the estimated effects of industry segregation rise for the more disaggregated occupation breakdowns. Finally, the contribution of occupational segregation varies significantly across the columns; even the sign changes across the columns, so that in the last two columns, occupational segregation actually narrows the wage gap between men and women by 16 to 30%. The change in the estimated effects of the female dummy variable does vary across columns, decreasing 2.0 percentage points in column (1) but increasing 0.4 and 3.4 percentage points in columns (2) and (3). In every model, however, there is a large role for individual sex even after accounting for the percent female within occupation, industry, establishment, and job cell. The female dummy variable accounts for at least 28% of the wage gap.

#### *Eliminating Measurement Error*

An approach that eliminates any role of measurement error in the percent female variables is to include a full set of job-cell dummy variables, which capture occupation, industry, establishment, and of course job-cell fixed effects. In addition to eliminating measurement error, this specification has the benefit that, unlike standard wage decompositions, it does not impose a particular parametric structure on the way segregation affects sex wage differences. The cost of this procedure is that we sacrifice estimates of the effects of sex segregation along each of these dimensions. These estimates are reported in Table 10. Because job-cell dummy variables are included, the estimated effects of the individual's sex in these specifications can be thought of as within-job-cell sex differences in wages. Up to this point, we have been reluctant to use this label in describing the effect of an individual's sex; however, because the results from this less parametric specification turn out to be quite similar to those using the percent-female variables, from this point on we use this more transparent expression to describe the sex difference in wages attributable to an individual's sex.

In the first two panels, estimates are first reported for wage equation specifications including no

controls, and then including the human capital controls. Compared with the corresponding estimates in Table 5, the estimated effect of the individual's sex is actually larger in Table 10 whatever the level of occupational detail.<sup>26</sup> For example, in Table 5 using the most-aggregated occupations, the estimated coefficient of the female dummy variable is  $-.132$ ; the comparable estimate in Table 10 is  $-.170$ . Interestingly, however, the decline in this estimated coefficient as more-detailed occupations are used is similar in Table 10 (from  $-.170$  to  $-.130$  using the most-detailed occupations) to the decline in Table 5 (from  $-.132$  to  $-.091$ ). As the last row of the table reports, the ratio of the estimated within-job-cell sex difference using the fixed effects vs. the percent-female variables is lower when we use job cells with at least 10 workers relative to the full sample, which is consistent with some decrease in measurement error bias as we restrict the sample to larger job cells.<sup>27</sup> Regardless, no matter what specification one looks at it in Table 10, there is no evidence that measurement error in the earlier tables is biasing the estimated within-job-cell sex difference in wages away from zero, leading to a spurious negative

---

<sup>26</sup>While it is impossible to sign a priori the effect of measurement error on the female dummy variable, one's intuition might be to expect the coefficient on the female dummy variable to be go down in Table 10 relative to the earlier tables since, in those tables, the female dummy is positively correlated with the mismeasured percent female in establishment and job-cell variables. Exactly the opposite occurs.

<sup>27</sup>We also used this specification to verify that the relationship we estimate between wages and sex is actually driven by the rate of pay. Recall that our wage variable is a constructed wage, with hours and weeks in the denominator. In principle, it is possible that it is not rates of pay on a weekly, bi-weekly, monthly, etc., basis that differ by sex within job cells, but rather that it is weeks or hours that differ. For women to have lower constructed wages within job cells, however, it would have to be the case that their weeks or hours were higher within job cells, which seems unlikely. To check this, we estimated specifications for log weeks and for log hours with a dummy variable for female and job-cell fixed effects as control variables. For the different levels of occupational disaggregation, the estimated coefficient on the female dummy variable in the weeks regression ranged from  $-.005$  to  $-.007$ , while in the hours regression it ranged from  $-.030$  to  $-.044$ . These negative coefficients imply that dividing through by weeks and hours would tend to make constructed wages look, if anything, more equal by sex within job cell than the reported rates of pay.

estimate.<sup>28</sup> Moreover, the results in Table 10 make it clear that the estimated within-job-cell sex difference in Table 5 is not attributable to the functional form used to estimate the impact of segregation on wages.

### *The Effects of Using Separate Regressions Estimated by Sex*

To this point, we have followed Groshen (1991) in using a single wage regression for men and women, with a dummy variable for women. A more flexible decomposition procedure is to use separate wage regressions for men and women, following Oaxaca (1973). In this case, using 'f' and 'm' subscripts to indicate coefficients corresponding to women and men, we can decompose the difference in average log wages between women and men as follows (dropping the variable  $F_p$  used in equation (1)):

$$\begin{aligned}
 (3) \quad w_f' - w_m' = & \{ ( \alpha_f' (OCC\%F_f - OCC\%F_m) + \beta_f' (IND\%F_f - IND\%F_m) + \gamma_f' (EST\%F_f - EST\%F_m) \\
 & + 2_f' (JOB\%F_f - JOB\%F_m) + (X_f - X_m)M_f' \} \\
 & + \{ ( \alpha_f' - \alpha_m' ) + ( \beta_f' - \beta_m' ) OCC\%F_m + ( \gamma_f' - \gamma_m' ) IND\%F_m + ( \delta_f' - \delta_m' ) EST\%F_m \\
 & + ( 2_f' - 2_m' ) JOB\%F_m + X_m (M_f' - M_m' ) \} .
 \end{aligned}$$

In this decomposition, the first term in curly brackets gives the component of the wage gap attributable to differences in characteristics of individuals and jobs, while the second gives the component attributable to differences in coefficients. In the usual application of this decomposition, the second term is interpreted as capturing discrimination, although once one starts to include job characteristics (such as the percent female in the occupation), one could argue that this interpretation breaks down because sex

---

<sup>28</sup>It is also the case that the standard errors hardly increase when we use fixed job-cell effects. In the specifications with the percent-female variables, the standard errors were potentially understated because there were multiple observations on the same establishment, job cell, etc. However, the fact that the standard errors are scarcely larger with fixed job-cell effects indicates that this understatement was trivial.

segregation may reflect discrimination.<sup>29</sup> As explained in Neumark (1988), the decomposition in equation (3) uses the female wage structure as the "no-discrimination" wage structure, which is why mean differences in characteristics are evaluated at the female coefficients. Alternatively, one can use the male (or some other) wage structure, in which case the coefficients at which the mean differences are evaluated would have 'm' subscripts, and the means at which the differences in coefficients are evaluated would in turn have 'f' subscripts.

In terms of the decompositions used in this paper, each of the first four terms in equation (3) gives the component of the wage differential attributable to the corresponding dimension of sex segregation. For example,  $(\alpha_f' (OCC\%F_f - OCC\%F_m))$  measures the component attributable to occupational segregation. In contrast, the entire second term involving the differences in intercepts and coefficients measures the component attributable to the individual's sex.

Table 11 reports results using this less-restrictive decomposition. We restrict attention to the two most-detailed occupational classifications in the Decennial Census. The first thing to notice is that the estimated coefficients on the percent-female variables differ sharply by sex. The percent female in the occupation has a negative and significant effect on men's wages but a small positive and significant effect on women's wages. The percent female in an industry has a much larger negative impact on men's wages than on women's wages, while, conversely, the percent female in the establishment and job cell have much stronger negative influences on women's wages. Not surprisingly, then, the decompositions will be markedly different depending on the choice of the no-discrimination wage structure, as reported further down in the table.

The best comparison is between the results reported in columns (2) and (3) of Table 5, and those reported in the lower panel of Table 11. This comparison indicates that, using the female wage structure

---

<sup>29</sup>In fact, there are also many reasons why differences in coefficients may not reflect discrimination, such as the apparent productivity-enhancing effects of marriage for men (Korenman and Neumark, 1992).

(reported following the slash), the shares of the wage gap attributable to each dimension of segregation and to within-job-cell sex differences are roughly similar; the only notable difference is that in Table 11 segregation by job cell has a larger effect. Using the male wage structure (reported before the slash), a considerably larger share (about 50%) of the wage gap arises within job cells, reflecting the weaker estimated effects of establishment and job-cell segregation for men. Although the apparent differences in the effects of segregation for men and women are of independent interest and bear further exploration, for the purposes of this paper the important point is that using separate wage equations for men and women certainly does not eliminate the significant role of within-job-cell sex differences in wages. Compared with single-equation estimates, these differences are as large using the female wage structure, and much larger using the male wage structure.

#### V. Comparison with Groshen's Estimates

To this point, we have presented numerous analyses of the NWECD data set. The findings indicate that within-job-cell sex differences in wages as well as job-cell segregation account for large shares of the sex wage gap, and that occupational segregation is relatively unimportant. The findings regarding the within-job-cell sex differences and occupational segregation contrast sharply with the findings reported by Groshen (1991).<sup>30</sup> She studies five specific industries in both manufacturing and non-manufacturing, using data from the 1970s and 1980s drawn from Bureau of Labor Statistics IWS. Groshen finds that within-job-cell sex differences account for only -1.0 to 6.6% of the wage gap. In contrast, the effect of job-cell segregation ranges from -2.7 to 32.5% of the wage gap, with the

---

<sup>30</sup>The result on occupational segregation may appear to contradict a rather extensive literature documenting the negative relationship between wages and the percent female (e.g., Macpherson and Hirsch, 1995; Sorenson, 1990). However, most of this work does not include information on other components of segregation. When we estimate the models in Table 5 omitting the percent female in industry, establishment, and job cell, we find strong negative effects of the percent female by occupation, which accounts for 6 to 21% of the sex wage gap, with the upper end of the range corresponding to the more-detailed occupational classification.

percentage above 20 for three of the five industries. The percent female in the occupation accounts for the largest share, ranging from 40.6 to 74.8%. The estimated contributions of job-cell segregation are not very different from ours, but Groshen's estimated contributions of occupational segregation and within-job-cell sex differences contrast strongly with our results.<sup>31</sup> In this section, we discuss a number of possible explanations of these differences.

### *Differences in Control Variables*

Because of data limitations in the IWS, Groshen does not have any human capital or other controls in her equations. The estimates in Tables 4-6, excluding and including controls, indicate that this does not explain the differences between the two sets of results. In particular, the inclusion of human capital and other controls in the NWECD data results in a remaining within-job-cell sex difference in wages that is smaller than the difference in the absence of controls. If we compare Groshen's results with estimates using the NWECD without these other controls (as in columns (2)-(5) of Table 4), the contrast of our results with hers is even sharper. To ensure more direct comparability of the results from the two data sets, in this section we report results excluding these other controls.

### *The Effects of the Particular Industries Studied*

The most obvious difference between our NWECD data set and the IWS data that Groshen studies is the scope of industries studied. The results we have presented using the NWECD are based

---

<sup>31</sup>Our finding that sex segregation within occupation accounts for a relatively small proportion of the male/female wage gap is consistent with the results reported by Macpherson and Hirsch (1995). These authors, using matched longitudinal data on workers from the Current Population Survey, find that once they control for unobserved worker characteristics the estimated effect of percent female within occupation declines substantially. In fact our estimated coefficient of -.038 (in column (3) of Table 5) is in the range of estimates they report in their Table 9. Since Macpherson and Hirsch only have data on workers, part of what is captured by controlling for unmeasured worker characteristics is the effect of percent female in a worker's establishment and job cell, which we find to be considerably more important than occupational segregation, yet which are of course positively correlated with percent female in the occupation. Although we cannot compare our results on the role of these other factors with Macpherson and Hirsch, their results do provide some support for the accuracy of our findings.

on a large sample consisting of a broad (nearly complete) array of industries, while Groshen's work is restricted to five specific industries: Miscellaneous Plastics Products, Nonelectrical Machinery, Banking, Life Insurance, and Computer and Data Processing. To examine whether her results are specific to the industries she studied, we attempted to draw subsamples in our data set for these specific industries. We were able to obtain samples large enough to study for the first three of these industries.<sup>32</sup> In addition to restricting the analysis of the NWECD data to these industries, we also restrict attention to the classes of occupations covered by the IWS. For Plastics and Machinery, these include production and maintenance occupations, while for Banking these include nonsupervisory occupations.

A direct comparison of the results for similar workers in Miscellaneous Plastics Products, Nonelectrical Machinery, and Banking is provided in Table 12. We present results using the NWECD with each level of occupational disaggregation in columns (1)-(3), and Groshen's results using the IWS in column (4).<sup>33</sup> Turning first to the segregation estimates, as in the full sample, job-cell segregation plays a fairly important role in the NWECD data. In each industry, the estimated contribution of job-cell segregation to the sex wage gap is similar to Groshen's estimates, especially with the more-detailed occupational classification. For example, using the most-detailed classification in the NWECD data, in column (3), the estimated share for Plastics is .342 in the NWECD data, vs. .325 in the IWS. In Nonelectrical Machinery the corresponding estimates are .254 vs. .124, and in Banking .208 vs. .218.

---

<sup>32</sup>In contrast, we had data on only 464 workers in 70 establishments in Computer and Data Processing, and fewer than 75 workers and fewer than 10 establishments in Life Insurance.

<sup>33</sup>The raw sex wage gaps are roughly comparable in the two data sources for the two manufacturing industries, but not banking, although always smaller in the IWS. In the NWECD the estimated coefficients of the sex dummy variable when it is the only included variable are -.32 for Plastics, -.40 for Nonelectrical Machinery, and -.56 for Banking. The corresponding numbers for the IWS reported in Table 1 of Groshen's paper are -.24, -.30, and -.37.



Similarly, the estimated contributions of establishment segregation are similar for the two data sources, with estimates in the NWECD and the IWS (again, comparing columns (3) and (4)) of .170 vs. .121 in Plastics, .322 vs. .480 in Nonelectrical Machinery, and .038 vs. .048 in Banking.

The differences between the two data sets remain sharp, however, regarding the roles of occupational segregation and the individual's sex (or within-job-cell sex differences in wages). In all three industries the IWS data indicate a much more prominent role for occupational segregation. In the two manufacturing industries, Groshen's estimates attribute 41 to 49% of the wage gap to occupational segregation, compared with nine to 22% in the NWECD data. In Banking, Groshen's estimates indicate an even larger role for occupational segregation, contributing 71% of the sex gap in wages, compared with 29% in the NWECD data. Most importantly, in each of the three industries we obtain the same result as in the full sample that the sex gap within job cells accounts for a large share of the wage gap; using the most-detailed occupational classification, the individual's sex accounts for 26 to 47% of the sex wage gap, compared with one to seven percent in the IWS data. Thus, even looking at the industries (and occupations) Groshen studies, our data assign a less prominent role to occupational segregation, and--in sharpest contrast to her estimates--a prominent role to sex wage gaps within establishments and occupations.<sup>34</sup>

### *Differences in Sample Periods*

One possible explanation is that our data are from a later period than the IWS, and it is

---

<sup>34</sup>Because her occupation definitions are industry-specific, Groshen computed percent female in the occupation separately for each industry studied, which contrasts with the procedure in more-standard data sets of using economy-wide estimates to study the role of occupational segregation. In contrast, we use the full SDF to compute the percent female. However, we reestimated the models in Table 12 using a definition of the percent female based only on the specific industry. This had little or no impact on the contribution of the individual's sex, although it did slightly raise the contribution of occupational segregation, and slightly lower the contribution of job-cell segregation.

conceivable that sex differences in wages that remain after accounting for each dimension of sex segregation widened over the 1980s. While we have no way to evaluate this possibility, we are skeptical that relative contributions of different components of the wage gap changed so much in such a short period. We know, of course, that wage differences between and within skill groups did widen over this period. However, given that the overall sex gap in wages did not widen over this period (Blau, 1998), we see no reason to believe that it widened within job cells. We also know that occupational segregation declined over this period (Blau, 1998), so it is possible that the contribution of occupational segregation fell, but there is no obvious reason why the wage gap previously attributable to occupational segregation would now all "load" on to the individual's sex. Finally, the differences between our results and Groshen's are as sharp or sharper for the Nonelectrical Machinery and Banking industries compared with Miscellaneous Plastics, despite the fact that the data from the first two industries come from the 1980s, while the data for the third come from the 1970s. Overall, then, we are very doubtful that differences in the periods studied explain the differences in results.

#### *Differences in Occupational Classification and Coverage*

Although in Table 12 we restrict attention to the same broad occupations that Groshen studies, there remain differences in the occupational classification. In particular, the occupational classifications available in the IWS are highly detailed, and specific to the five narrow industries Groshen studies.<sup>35</sup> This raises the possibility that if we could somehow use occupational classifications in the NWECD that are closer to those used by Groshen in the IWS, our results would coincide with hers.<sup>36</sup> However, the

---

<sup>35</sup>As an example, in the Miscellaneous Plastics Products industry, Groshen cites three different jobs on blow-molding machines as "operate," "set-up," and "set-up and operate."

<sup>36</sup>Blau's (1977) occupational classifications are also somewhat narrower than ours, primarily because she has some distinctions based on job levels within occupations.

results in Table 5 show that going from 13 to 491 occupations in the NWECD makes little substantive difference, which makes us skeptical that the level of occupational detail explains the differences in results. Furthermore, although the IWS occupational classification may differ from ours, it is not clear that the IWS uses a greater degree of occupational disaggregation within each industry. In the top row of each panel in Table 12 we report the number of occupations represented in the data set when we use each level of occupational disaggregation; in columns (1)-(3) this is the number following the slash, whereas the number preceding the slash is the total number of occupations on which the classification is based. These numbers reveal that using the most-detailed Census occupations (in column (3)), we actually have more occupations represented in each of the three industries than the IWS.

On the other hand, it may be that features of the occupational classification other than the level of disaggregation are relevant. Presumably the most relevant one is the extent to which the occupational classification is based on sex-segregated occupations. For example, if the Census occupations are less sex segregated, and hence combine occupations or jobs based on the IWS classifications that are segregated by sex, then in the NWECD we may find more evidence of a within-job-cell sex gap in wages after accounting for segregation.

It appears, however, that the degree of sex segregation in the detailed Census occupations is no lower than in Groshen's data. In particular, Table 13 compares the mean difference in the percent female in job cells in the two data sets, again for the NWECD using the most-detailed occupational classifications. As the table shows, once we use relatively detailed Census occupational classifications, sex segregation in the NWECD is about the same as in the IWS in the two manufacturing industries,

and higher in Banking.<sup>37</sup>

Another possibility is that the results are simply sensitive to the precise occupations covered within each industry. It turns out that the IWS often covers a narrow set of occupations within each industry, among the production and maintenance workers (in Plastics and Nonelectrical Machinery) and nonsupervisory workers (in Banking). To address this question, we also constructed subsamples for each of these industries in which we restrict attention to the subset of Census occupations that can be matched reasonably well with the main occupational categories surveyed by the IWS. We can do this because the occupational descriptions provided in the IWS documentation correspond closely to those in the SOC, and we can then map the SOC codes into Census codes.<sup>38</sup> Note that this alternative approach does not yield greater occupational disaggregation in the NWECD. All it yields is an analysis similar to the one we have done, but for the specific occupations covered in the IWS.

The results using these subsets of occupations are reported in Table 14. The results for the IWS occupations are reported in column (1). As shown in the first row of each panel, there are now, of course, fewer occupations covered. A comparison of the estimates in column (1) with those in column (2), which are for all of the occupations within the broad categories of production and maintenance or nonsupervisory workers, indicates that the estimates are not very sensitive to restricting the sample to the subset of occupations covered in the IWS. The estimated effects of each type of segregation are similar in the two samples. Furthermore, the estimated effect of the individual's sex remains large for each industry in the subset of IWS occupations, with its estimated contribution to the wage gap similar in this subset and the broader set of occupations. Thus, differences in occupational coverage do not explain the

---

<sup>37</sup>The results of this comparison were similar if we restricted attention to the occupations that match more explicitly those covered by the IWS, discussed below.

<sup>38</sup>This refers to the main occupational categories surveyed in the IWS. The Census occupations are too aggregated to be matched to the occupation sub-categories in the IWS (which distinguish, for example, by job level).

differences between our results and Groshen's.

The only possibility is that within the particular occupations considered in column (1) of Table 14, the degree of occupational disaggregation matters; this is the same question we considered earlier with respect to the broader set of occupations in the NWECD. Although it is true that for these particular occupations Groshen's occupational classification is more disaggregated, as indicated by the greater number of occupations in column (3) than in column (1) of Table 14, the essential question is whether the aggregation in the NWECD masks sex segregation. To examine this question, we computed the measure of sex segregation based on the sex difference in the mean percent female in the job cell, for the specific occupations corresponding to those covered by the IWS (as we did for the broader set of NWECD occupations in Table 13). As in Table 13, the data again indicate that the aggregation in the NWECD does not mask sex segregation relative to Groshen's classification. The degree of sex segregation is now greater in the NWECD for both Plastics (.775 vs. .761) and Banking (.681 vs. .505), and is only a bit smaller in Nonelectrical Machinery (.596 vs. .646). Thus, we conclude that neither differences in occupational classification nor in occupational coverage appear to explain the differences between our results and Groshen's.

### *Measurement Error*

As discussed earlier, measurement error may influence some of our results relative to Groshen's. In particular, sampling error in the estimates of the percent female in the establishment and job cell may be quite severe. Of course, Groshen also does not have data on all workers in each establishment, but rather just those in a subset of occupations; specifically, the IWS excluded executive, professional, and supervisory workers. Thus, her estimates of the percent female in the establishment are also prone to sampling error, and may also be influenced by the non-random selection of which occupations were surveyed. Similarly, her estimates of percent female in the occupation are based on a

small subset of establishments, also non-random, and hence are prone to error, whereas our estimates of occupational segregation are based on the full SDF, so that measurement error bias is likely minimal.<sup>39</sup> In addition to these considerations, the analysis restricting attention to large job cells (Table 9) and using fixed effects for occupation, industry, establishment, and job cell (Table 10), suggested that measurement error in the percent female variables was not responsible for finding a significant negative sex gap in wages within job cells.

Finally, the direct comparison of estimates on an industry-by-industry basis also permits us to make some more definitive statements about the possible explanations of the differences between our results and those of Groshen. The variables in the NWECD for which measurement error is potentially serious are the percent female in the establishment and the percent female in the job cell. To address questions of measurement error, it is more useful to look at coefficient estimates than the decomposition results, as the latter are also multiplied by the mean differences in the segregation measure. These results are reported in Table 15. We focus on the comparison between the two most-detailed occupational classifications in the NWECD, and Groshen's estimates.

The variable that is most prone to measurement error in the NWECD is the percent female in the job cell. However, the estimates in Table 15 indicate that the estimated coefficients of this variable are larger in the NWECD than in the IWS for all three industries. This indicates that measurement error in the NWECD, which should bias the estimated coefficient of percent female in the job cell downward (toward zero), is not responsible for the differences in results. Reinforcing this conclusion, the percent female in the establishment is also prone to measurement error in the NWECD, yet the estimated

---

<sup>39</sup>In fact, in the discussion of the IWS, the BLS Handbook of Methods states that "estimates of the number of workers in a given occupation are subject to considerable sampling error, due to the wide variation among establishments in the proportion of workers found in individual occupations."

coefficient of this variable is higher in our sample than in the IWS data for the two manufacturing industries, although not for Banking. Based on this evidence, we conclude that although measurement error may bias some of the estimates we obtain using the NWECD data, it does not explain the differences between our results and Groshen's.

### *Data Collection*

To this point, we have found that issues regarding measurement, occupational classification, industries studied, etc., do not appear to explain the differences between our results and Groshen's. We next turn to some issues regarding the role that data collection may play.

In the IWS, within each industry establishments were chosen using a random probability sample with over-weighting of large establishments. Selected establishments reported information on the number of workers of each sex working in nonsupervisory production jobs and nonsupervisory nonproduction jobs (office work).<sup>40</sup> No data on the sex composition of supervisory workers were collected. Payroll records ideally were then used to collect data on the number of workers of each sex working in a specific (non-supervisory) occupation at each earnings level. Data were only reported for a selection of occupations that were chosen on an industry-by-industry basis partially to reflect the types of activities performed by workers in an industry, but also to reflect their "prevalence in the industry, definiteness and clarity of duties, and importance as reference points in collective bargaining." (U.S. Bureau of Labor Statistics, 1976).

The IWS data have a number of strengths and weaknesses. Because the data on the sex composition of workers in each job cell ideally were generated using payroll data, there should be no

---

<sup>40</sup>The survey, BLS 2751A, asks the establishment to report the "approximate total employment" in the establishment.

sampling error. In contrast, our NWECD data do contain this kind of sampling error. On the other hand, as noted earlier, the IWS likely has some measurement error in the percent female in the establishment and the occupation.

Perhaps the biggest drawback to the IWS relative to the NWECD, however, is that all the data in the IWS come from surveyed establishments. While it is hard to gauge how accurately establishments were able to or required to provide detailed occupational and earnings data by sex, it is clear that it is not in any establishment's best interest to provide data suggesting that within narrowly-defined occupations there was a pay gap between men and women.<sup>41</sup> It might therefore not be surprising to find that there is little evidence of a within-job-cell sex wage gap in the IWS. In contrast, the wage and sex data in the NWECD come from data reported by the worker in the SDF, where the worker has no incentive to misreport earnings or occupation in a way that is systematically correlated with sex. While this is only a possible explanation for the differences between the results from the two data sets, we have not been able to verify that the data collection effort accessed employers' payroll records in such a way as to ensure that employers did not misreport individuals' wages to mask within-job-cell sex differences. For more details on how the IWS data on earnings and sex-mix within occupations were collected, see the Appendix.

Finally, to this point we have only offered a conjecture that the IWS data lead to masking of within-job-cell sex wage gaps. Within-job-cell sex wage gaps could be masked either by misreporting of wages, or by misreporting of the sex composition of occupations. Given that we do not find more segregation in the IWS data than in the NWECD data, we doubt that misreporting of the sex composition of occupations in the IWS--specifically, the reporting of more segregation than is actually the case--is

---

<sup>41</sup>The Equal Pay Act places the burden of proof on the employer to show that unequal pay for equal work is based on a factor other than sex, such as seniority.



occurring. However, we do note that in each of the three industries that we can compare, the raw sex wage gap is smaller in Groshen's data than in our data (see footnote 33). As noted earlier, in Groshen's data, the gap is -.24 in Plastics, -.30 in Nonelectrical Machinery, and -.37 in Banking; in contrast, the corresponding gaps are -.32, -.39, and -.56 in our data. Thus, despite that fact that Groshen's data come from an earlier period when the sex wage gaps would have been expected to be larger, they are smaller by .08 to .19 in her data. This is consistent, at least, with underreporting of sex wage gaps in the IWS.<sup>42</sup>

We do not have direct and compelling evidence to suggest that the IWS be disregarded for the purpose of decomposing the sex wage gap. However, given the distinctly different findings produced using the NWECD, and the potential problems with the IWS, our findings should, at a minimum, call into question the previously-held belief based on the IWS data that occupation, establishment, and job-cell segregation can account for essentially all of the male/female wage gap that is not explained by other individual-level characteristics.

## VI. Policy Implications

The policy implications of our findings are very different than those drawn from the earlier research. Our results suggest that identifying and eliminating the sources of within-occupation, within-establishment wage differences between men and women can play a large role in reducing wage differences between men and women. In particular, if within the narrowly-defined occupations that we study, the jobs performed by men and women require substantially equal skill, effort, responsibility, and working conditions, yet wages differ by sex, then enforcement of the Equal Pay Act has a fundamental role to play in closing the wage gap between men and women. In contrast, if segregation along various dimensions accounted for most of the sex wage gap, then policies along the lines of comparable worth,

---

<sup>42</sup>We thank Bob Lalonde for suggesting this comparison.

equal opportunities in employment and promotion, and affirmative action would be central to any further closing of this gap, and stronger equal pay provisions would serve no purpose.<sup>43</sup>

Indeed, there is substantial interest in equal pay issues within segments of the federal government. Similar bills introduced recently in the U.S. Senate and U.S. House of Representatives call for strengthening of the Equal Pay Act to allow for, among other things, compensatory and punitive damages against employers violating the Act.<sup>44,45</sup> In addition, the Senate version of the bill instructs the Secretary of Labor to "develop guidelines to enable employers to evaluate job categories based on objective criteria such as educational requirements, skill requirements, independence, working conditions and responsibility" which employers can use to help ensure that they are in compliance with the law. The House version instructs the Secretary of Labor to issue "voluntary pay guidelines for the relative pay ranges of a selection of male- and female-dominated widely held occupations." Finally, the House version of the bill requires employers with 100 or more employees to submit reports to the EEOC containing information from payroll records on wages by race, sex, and national origin of employees. The Clinton administration recently endorsed the legislation.<sup>46</sup> While not addressed to the specifics of these proposals, our findings suggest that stronger equal pay legislation may further reduce the wage gap between men and women.

---

<sup>43</sup>For example, referring to the role of equal pay provisions in eliminating within-job-cell sex differences in wages, Groshen concludes from her results that "either the provisions have succeeded, or this component was never the main source of the gap" (p. 470).

<sup>44</sup>The Senate Bill is S.71, introduced by Senator Thomas Daschle on January 21, 1997. The House Bill is H.R. 2023, introduced by Rep. Rosa DeLauro on June 24, 1997.

<sup>45</sup>Attorneys' fees can currently be awarded.

<sup>46</sup>See "Vice President Gore Announces Support for Legislation on Equal Pay," White House Press Release, April 2, 1998.

## VII. Conclusions

We assembled a large matched employer-employee data set covering essentially all industries and occupations across all regions of the U.S. We use this data set to re-examine the question of the relative contributions to the overall sex gap in wages of sex segregation vs. wage differences by sex within occupation, industry, establishment, and occupation-establishment cells. This is especially important given that earlier research on this topic relied on data sets that covered only a narrow range of industries, occupations, or regions.

Our results indicate that, although a sizable fraction of the sex gap in wages is accounted for by the segregation of women into lower-paying occupations, industries, establishments, and occupations within establishments, a substantial part of this gap remains attributable to the individual's sex. Overall, our estimates indicate that from about one-quarter to one-half of the sex wage gap takes the form of wage differences between men and women within narrowly-defined occupations within establishments. These findings contrast sharply with the conclusions of previous research (especially Groshen, 1991) indicating that sex segregation accounted for essentially all of the sex wage gap. We explore numerous possible explanations for these different findings, but we find that even restricting attention to the narrow industries studied in the earlier research, our data still show important differences in wages paid to men and women within occupations and establishments.

We do not attempt in this paper to determine the underlying forces that cause men and women to have different wages within narrowly-defined occupations in the same establishments. Our results simply suggest that there is still research to be done in order to identify these forces. In particular, our results leave open the possibility that within narrowly-defined occupations and establishments, men and women are performing essentially the same job but are not being paid equally--a violation of the Equal

Pay Act. Further research into the sources of within-establishment within-occupation sex wage differences is therefore much more important than previously thought.

## Appendix: Issues in the Collection of Earnings and Sex-Mix Data by Occupation in the IWS

Ideally, BLS field representatives collected data for the IWS on earnings and the sex-mix within occupations using the payroll records of surveyed establishments. Data were only reported for a selection of occupations that were chosen on an industry-by-industry basis partially to reflect the types of activities performed by workers in an industry, but also to reflect their "prevalence in the industry, definiteness, and clarity of duties, and importance as reference points in collective bargaining."

When payroll records were utilized for data collection, it is not entirely clear how field economists verified or coded wages. The form used to record the data (BLS 2753G, reproduced in the BLS Handbook of Methods) has the following headings: (i) occupation and grade, (ii) occupational code, (iii) sex, (iv) method of pay, (v) number of workers, (vi) hours, and (vii) salary, rate, or earnings. There is some ambiguity regarding how the field economist would record data on individual men and women. It does not seem likely that a record was filled out for each worker, both because the number of records could be prohibitively large, and because the entry for "number of workers" would not make sense. Yet the Handbook indicates that this form was used "in recording occupation, sex, ..., and pay rate or earnings for each worker studied" (p. 138). Alternatively, it is possible that the field economist filled out a separate line for workers of each sex within an occupation,<sup>47</sup> but it also appears possible that she would record a single wage for the occupation, indicating the percent female. Unfortunately, however, we have been unable to track down instructions that may have been issued to field economists to better understand how the IWS data were meant to be collected, or original copies of

---

<sup>47</sup>It seems somewhat implausible to us to think that in years subsequent to the passage of the Equal Employment Opportunity Act of 1972 companies worked with government field economists to prepare reports indicating different wage rates for men and women doing the same job, despite a pledge of confidentiality from the BLS.

the filled-out surveys to determine how the data were actually collected.

The BLS documentation for the IWS also suggests that in at least some cases companies generated the reports for the surveys (U.S. Bureau of Labor Statistics, 1976). The Handbook (p. 138) reads:

"Bureau field economists collect data by personal visit to each of the sample establishments. Job functions and factors in the establishment are carefully compared with those included in the Bureau job definitions. The job matching may involve review of records such as pay structure plans and organization charts, company position descriptions, interviews with appropriate officials, and, on occasion, observation of jobs within plants. A satisfactory completion of job matching permits acceptance of company-prepared reports where this procedure is preferred by the respondent. Generally, however, the field economist secures wages or salary rates ... from payroll or other records ..."

It is hard to imagine that self-generated reports by companies would contain obvious evidence of violation of the Equal Pay Act. We were able to find no information on how many companies generated their own reports for the BLS.

## References

Bielby, William, and James Baron. 1984. "A Woman's Place is with Other Women: Sex Segregation Within Organizations." In Barbara Reskin, ed., Sex Segregation in the Workplace: Trends, Explanations, Remedies (Washington, DC: National Academy Press), pp. 27-55.

Blau, Francine D. 1977. Equal Pay in the Office (Lexington, MA: D.C. Heath and Company).

Blau, Francine D. 1998. "Trends in the Well-Being of American Women, 1970-1995." Journal of Economic Literature, Vol. 36, No. 1, March, pp. 112-65.

Brown, Charles, and James L. Medoff. 1989. "The Employer Size Wage Effect." Journal of Political Economy, Vol. 97, No. 5, October, pp. 1027-59.

Carrington, William J., and Kenneth R. Troske. "Sex Segregation in U.S. Manufacturing." Industrial and Labor Relations Review, Vol. 51, No. 3, April, pp. 445-64.

Fields, Judith, and Edward N. Wolff. 1995. "Interindustry Wage Differentials and the Gender Wage Gap." Industrial and Labor Relations Review, Vol. 49, No. 1, October, pp. 105-120.

Gold, Michael Evan. 1993. An Introduction to the Law of Employment Discrimination, ILR Bulletin 68 (Ithaca, NY: ILR Press).

Groshen, Erica L. 1991. "The Structure of the Female/Male Wage Differential: Is it Who You Are, What You Do, or Where You Work?" Journal of Human Resources, Vol. 26, No. 3, Summer, pp. 457-72.

Jacobsen, Joyce P. 1994. "Trends in Workforce Segregation 1960-1990." Social Science Quarterly, Vol. 75, No. 1, March, pp. 204-11.

Johnson, George, and Gary Solon. 1986. "Estimates of the Direct Effects of Comparable Worth Policy." American Economic Review, Vol. 76, No. 5, December, pp. 1117-25.

Korenman, Sanders, and David Neumark. 1992. "Marriage, Motherhood, and Wages." Journal of Human Resources, Vol. 27, No. 2, Spring, pp. 233-55.

Korenman, Sanders, and David Neumark. 1991. "Does Marriage Really Make Men More Productive." Journal of Human Resources, Vol. 26, No. 2, Spring, pp. 282-307.

Macpherson, David A., and Barry T. Hirsch. 1995. "Wages and Gender Composition: Why Do Women's Jobs Pay Less?" Journal of Labor Economics, Vol. 13, No. 3, July, pp. 426-71.

Neumark, David. 1988. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination." Journal of Human Resources, Vol. 23, No. 3, Summer, pp. 279-95.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." International Economic Review, Vol. 14, No. 3, October, pp. 693-709.

Polachek, Solomon. 1975. "Differences in Post-School Investment as a Determinant of Market Wage Differentials." International Economic Review, Vol. 16, No. 2, pp. 451-70.

Sorenson, Elaine. 1990. "The Crowding Hypothesis and Comparable Worth." Journal of Human Resources, Vol. 25, No. 1, pp. 55-89.

Sorenson, Elaine. 1989. "Measuring the Pay Disparity Between Typically Female Occupations and Other Jobs: A Bivariate Selectivity Approach." Industrial and Labor Relations Review, Vol. 42, No. 4, July, pp. 624-39.

Stinson, John F. Jr. 1997. "New Data on Multiple Jobholding Available from the CPS." Monthly Labor Review, March, pp. 3-8.

Troske, Kenneth R. "The Worker-Establishment Characteristics Database." Forthcoming in John Haltiwanger, Marilyn Manser, and Robert Topel, eds., Labor Statistics Measurement Issues (Chicago: University of Chicago Press).

U.S. Bureau of Labor Statistics. 1976. BLS Handbook of Methods, Bulletin 1910.

U.S. Department of Commerce, Office of Federal Statistical Policy and Standards. 1980. The Standard Occupational Classification Manual (Washington, DC: U.S. Government Printing Office).



Table 1: Construction of NWECD

	Number of establishments (SSEL)	Number of workers (SDF)	Avg. number of matched workers	Avg. establishment employment from SSEL data	Avg. establishment earnings/worker from SSEL data	Avg. establishment earnings/worker from SDF data	Avg. earnings/ worker from SDF data
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
a. Establishments in SSEL with positive employment	5,593,379	...	...	21.10 (.45)	19,310.37 (10.98)	...	...
b. Establishments in unique industry-location cells	388,787	...	...	41.17 (2.30)	17,102.78 (28.97)	...	...
c. Workers among long-form respondents in Sample Detail File (SDF)	...	17,311,211	...	...	...	...	20,977.51 (6.75)
d. Respondents matched to establishments in b.	201,944	1,720,423	8.52 (.08)	66.46 (4.43)	18,123.98 (205.00)	17,094.00 (38.83)	20,831.89 (17.28)
e. Discard matches with imputed industry or location, or number of matched workers greater than SSEL employment, discard workers with zero or missing earnings, or working outside U.S.	156,332	1,117,424	7.15 (.07)	83.24 (5.72)	18,218.33 (181.34)	19,416.60 (47.81)	22,582.18 (21.21)
f. Exclude workers with hours < 30 or > 65, weeks in 1989 < 30, age < 18 or > 65, wage < \$2.50 or > \$500, public administration	129,021	845,036	6.55 (.07)	83.59 (6.70)	19,224.23 (218.75)	23,112.66 (53.66)	25,611.43 (23.88)
g. Discard establishments with employment < 25, number of matched workers < 5 percent of SSEL employment, and earnings outliers, final sample	32,931	637,718	19.37 (.26)	180.84 (2.68)	20,983.40 (63.64)	23,327.76 (71.73)	25,978.52 (26.58)

Standard errors of means are reported in parentheses.

Table 2: Descriptive Statistics for NWECD and Full SDF Sample, Full-Time Full-Year Workers

	<i>NWECD</i>		<i>SDF</i>	
	Mean (1)	Standard deviation (2)	Mean (3)	Standard deviation (4)
Annual earnings	25978.52	21227.38	27259.79	29275.13
Log hourly wage	2.349	.539	2.336	.600
<i><u>Demographics:</u></i>				
Female	.468		.428	
Age	39.589	10.979	38.440	11.272
Black	.070		.077	
Ever married	.855		.803	
Number of children ever born (women)	1.841	1.622	1.603	1.583
High school degree	.369		.325	
Some college	.191		.216	
Associate's degree	.095		.078	
Bachelor's degree	.128		.159	
Advanced degree	.083		.085	
<i><u>Location:</u></i>				
MSA	.575		.764	
New England	.044		.056	
Mid-Atlantic	.143		.153	
East North Central	.273		.195	
West North Central	.122		.092	
South Atlantic	.137		.165	
East South Central	.083		.059	
West South Central	.086		.095	
Mountain	.038		.050	
Pacific	.075		.134	
<i><u>Occupation:</u></i>				
Manager	.270		.281	
Support	.230		.306	
Service	.086		.088	
Farming	.004		.016	
Production	.145		.135	
Laborer	.265		.173	
<i><u>Industry:</u></i>				
Agriculture	.004		.018	
Mining	.006		.009	
Construction	.0001		.067	
Manufacturing	.517		.239	
Transportation	.062		.085	
Wholesale	.012		.052	
Retail	.039		.139	
FIRE	.007		.076	
Business services	.004		.043	
Personal services	.002		.023	
Entertainment services	.002		.010	
Professional services	.343		.238	

The sample size is 637,718 for the NWECD, and 10,830,247 for the SDF.

Table 3: Regressions for Log Wages for SDF and NWECD Workers

	<u>SDF workers</u>		<u>NWECD workers</u>	
	(1)	(2)	(3)	(4)
Female	-0.231 (0.001)	-0.174 (0.001)	-0.282 (0.001)	-0.166 (0.002)
Age	0.062 (0.0003)	0.048 (0.0003)	0.058 (0.0004)	0.043 (0.0003)
Age <sup>2</sup> /100	-0.060 (0.0004)	-0.045 (0.0003)	-0.055 (0.0004)	-0.040 (0.0004)
Number of children	-0.049 (0.002)	-0.032 (0.002)	-0.029 (0.002)	-0.023 (0.002)
Number of children $\times$ age	-0.0002 (0.00004)	-0.0002 (0.00003)	-0.0004 (0.00004)	-0.0002 (0.00003)
Black	-0.072 (0.002)	-0.031 (0.002)	-0.081 (0.002)	-0.042 (0.002)
Ever married	0.135 (0.001)	0.086 (0.001)	0.132 (0.002)	0.081 (0.002)
High school degree	0.178 (0.001)	0.104 (0.001)	0.156 (0.002)	0.098 (0.002)
Some college	0.284 (0.002)	0.156 (0.002)	0.254 (0.002)	0.143 (0.002)
Associate's degree	0.363 (0.002)	0.186 (0.002)	0.401 (0.002)	0.194 (0.002)
Bachelor's degree	0.568 (0.002)	0.335 (0.002)	0.542 (0.002)	0.325 (0.002)
Advanced degree	0.743 (0.002)	0.493 (0.002)	0.663 (0.002)	0.480 (0.003)
MSA	0.205 (0.001)	0.166 (0.001)	0.157 (0.001)	0.120 (0.001)
Industry controls	No	Yes	No	Yes
Occupation controls	No	Yes	No	Yes
R <sup>2</sup>	0.364	0.494	0.409	0.562
N	1,084,379	1,084,379	637,718	637,718

The regressions in columns (1) and (2) are based on a 10-percent random sample from the SDF. Standard errors of regression estimates are reported in parentheses. The industry and occupation controls include the full set of three-digit Census industry and occupation codes. Number of children refers to number of children ever born; this is asked of women in the Census, and is set to zero for men.

Table 4: Estimated Log Wage Differentials by Sex, and Percent Female in Occupation, Industry, Establishment, and Job Cell

	Coefficient estimate	Coefficient estimate	Mean difference, women - men	Absolute contribution to wage gap, (2) $\times$ (3)	Relative contribution to wage gap	Coefficient estimate	Relative contribution to wage gap
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-.375 (.001)	-.244 (.002)	1.00	-.244	.651	-.132 (.002)	.351
% female in occupation,	...	.180 (.004)	.180	.032	-.087	-.102 (.003)	.049
% female in industry	...	.122 (.006)	.248	.030	-.081	-.163 (.005)	.108
% female in establishment	...	-.188 (.005)	.338	-.064	.170	-.166 (.004)	.149
% female in job cell	...	-.243 (.004)	.536	-.130	.347	-.097 (.003)	.138
Age	...	...	-.312	...	...	.054 (.0004)	.045
Age squared $\times 100^{-2}$	...	...	-.243	...	...	-.051 (.0004)	-.033
Black	...	...	.023	...	...	-.076 (.002)	.005
Ever married	...	...	-.011	...	...	.120 (.002)	.004
Number of children	...	...	1.841	...	...	-.019 (.002)	.095
Number of children $\times$ age	...	...	81.590	...	...	-.0005 (.00004)	.103
High school degree	...	...	-.019	...	...	.159 (.002)	.008
Some college	...	...	-.001	...	...	.272 (.002)	.001
Associate's degree	...	...	.050	...	...	.450 (.002)	-.060
Bachelor's degree	...	...	.002	...	...	.592 (.002)	-.003
Advanced degree	...	...	-.011	...	...	.752 (.002)	.021
MSA	...	...	-.049	...	...	.146	.019

			(.001)
R <sup>2</sup>	.121	.140	.438

---

The sample size is 637,718. Standard errors of regression estimates are reported in parentheses. In this table, 13 occupational categories are used.

Table 5: Estimated Log Wage Differentials by Sex, and Percent Female in Occupation, Industry, Establishment, and Job Cell, Varying Degrees of Occupational Disaggregation

	(1)	(2)	(3)
Number of occupations	13	72	491
<i>Mean differences, women - men:</i>			
% female in occupation	.180	.340	.396
% female in job cell	.536	.657	.744
<i>Coefficient estimates:</i>			
Female	-.132 (.002)	-.104 (.002)	-.091 (.002)
% female in occupation	-.102 (.003)	-.011 (.003)	-.038 (.003)
% female in industry	-.163 (.005)	-.178 (.005)	-.164 (.005)
% female in establishment	-.166 (.004)	-.162 (.004)	-.187 (.004)
% female in job cell	-.097 (.003)	-.139 (.003)	-.119 (.003)
R <sup>2</sup>	.438	.438	.437
<i>Relative contribution to wage gap:</i>			
Female	.351	.278	.243
% female in occupation	.049	.010	.040
% female in industry	.108	.118	.109
% female in establishment	.149	.146	.169
% female in job cell	.138	.244	.236

All specifications include the same controls as in column (6) of Table 4. Column (1) reproduces results from columns (6) and (7) of Table 4. The sample size is 637,718. The residual category in the relative contribution to the wage gap is the contribution of the human capital controls.

Table 6: Estimated Log Wage Differentials by Sex, and Percent Female in Occupation, Industry, Establishment, and Job Cell, Varying Degrees of Occupational Disaggregation, with Establishment-Level Controls for Size and Multi-Unit Firms

	(1)	(2)	(3)
Number of occupations	13	72	491
<i>Mean differences, women - men:</i>			
% female in occupation	.180	.340	.396
% female in job cell	.536	.657	.744
<i>Coefficient estimates:</i>			
Female	-.135 (.002)	-.107 (.002)	-.094 (.002)
% female in occupation	-.095 (.003)	-.009 (.003)	-.042 (.003)
% female in industry	-.154 (.005)	-.168 (.005)	-.151 (.005)
% female in establishment	-.147 (.004)	-.145 (.004)	-.172 (.004)
% female in job cell	-.102 (.003)	-.143 (.003)	-.119 (.003)
R <sup>2</sup>	.457	.457	.456
<i>Relative contribution to wage gap:</i>			
Female	.361	.287	.252
% female in occupation	.046	.008	.044
% female in industry	.102	.111	.100
% female in establishment	.132	.131	.155
% female in job cell	.145	.250	.236

Dummy variables for establishment sizes 25-49, 50-99, 100-249, 250-499, and 500+ were included. Individual coefficient estimates are not reported, but indicated a significant positive monotone relationship of log wages with size. A control variable for whether the establishment belonged to a multi-unit firm was also included; this contributed .022-.023 to the wage gap. Aside from these differences, all specifications include the same controls as in column (6) of Table 4. The sample size is 637,718.



Table 7: Estimated Log Wage Differentials by Sex, and Percent Female in Occupation, Industry, Establishment, and Job Cell, Varying Degrees of Occupational Disaggregation, for Manufacturing and Non-Manufacturing Plants

<i>A. Manufacturing</i>			
	(1)	(2)	(3)
Number of occupations	13	72	491
<i>Mean differences, women - men:</i>			
% female in occupation	.112	.226	.287
% female in industry	.127	.127	.127
% female in establishment	.246	.246	.246
% female in job cell	.482	.632	.730
<i>Relative contribution to wage gap:</i>			
Female	.354	.284	.248
% female in occupation	.003	.049	.077
% female in industry	.095	.086	.083
% female in establishment	.092	.122	.133
% female in job cell	.172	.173	.174
<i>B. Non-Manufacturing</i>			
<i>Mean differences, women - men:</i>			
% female in occupation	.174	.344	.394
% female in industry	.205	.205	.205
% female in establishment	.275	.275	.275
% female in job cell	.480	.603	.698
<i>Relative contribution to wage gap:</i>			
Female	.339	.263	.233
% female in occupation	.148	-.003	.018
% female in industry	.097	.120	.115
% female in establishment	.091	.050	.070
% female in job cell	-.014	.237	.231

All specifications include the same controls as in column (6) of Table 4. The sample size is 329,905 in Panel A, and 307,813 in Panel B.

Table 8: Estimated Log Wage Differentials by Sex, and Percent Female in Occupation, Industry, Establishment, and Job Cell, Workers Under Median Age

	(1)	(2)	(3)
Number of occupations	13	72	491
<i>Mean differences, women - men:</i>			
% female in occupation	.177	.334	.386
% female in industry	.240	.240	.240
% female in establishment	.328	.328	.328
% female in job cell	.526	.649	.738
<i>Coefficient estimates:</i>			
Female	-.081 (.002)	-.065 (.003)	-.053 (.003)
% female in occupation	-.106 (.004)	.002 (.004)	-.025 (.004)
% female in industry	-.143 (.006)	-.164 (.006)	-.151 (.006)
% female in establishment	-.154 (.006)	-.155 (.006)	-.176 (.005)
% female in job cell	-.084 (.004)	-.114 (.004)	-.099 (.004)
R <sup>2</sup>	.398	.397	.397
<i>Relative contribution to wage gap:</i>			
Female	.308	.247	.200
% female in occupation	.071	-.002	.037
% female in industry	.130	.149	.137
% female in establishment	.192	.192	.218
% female in job cell	.167	.279	.276

All specifications include the same controls as in column (6) of Table 4. The sample size is 313,045. The raw wage gap for this sample is -.264. The median age is 39 for both men and women.

Table 9: Estimated Log Wage Differentials by Sex, and Percent Female in Occupation, Industry, Establishment, and Job Cell, Restricting Minimum Size of Job Cell to 10 Workers to Reduce Measurement Error

	(1)	(2)	(3)
Number of occupations	13	72	491
<i>Mean differences, women - men:</i>			
% female in occupation	.172	.357	.421
% female in industry	.266	.283	.292
% female in establishment	.338	.354	.364
% female in job cell	.439	.492	.535
<i>Coefficient estimates:</i>			
Female	-.122 (.002)	-.096 (.003)	-.091 (.004)
% female in occupation	-.160 (.005)	.151 (.008)	.227 (.010)
% female in industry	-.092 (.008)	-.309 (.012)	-.352 (.015)
% female in establishment	-.143 (.010)	-.011 (.012)	-.075 (.015)
% female in job cell	-.171 (.007)	-.290 (.008)	-.280 (.010)
R <sup>2</sup>	.373	.366	.367
N	448,305	298,227	201,176
Number of establishments	9,005	6,467	4,961
<i>Relative contribution to wage gap, restricted/unrestricted:</i>			
Female	.331/.351	.282/.278	.277/.243
% female in occupation	.075/.049	-.159/.010	-.291/.040
% female in industry	.067/.108	.258/.118	.312/.109
% female in establishment	.131/.149	.011/.146	.083/.169
% female in job cell	.203/.138	.421/.244	.455/.236

All specifications include the same controls as in column (6) of Table 4. In the bottom panel, the estimates following the '/' reproduce those in Table 5.

Table 10: Estimated Log Wage Differentials by Sex, Including Fixed Occupation, Industry, Establishment, and Job Cell Effects

	(1)	(2)	(3)
Number of occupations	13	72	491
<i>No controls, full sample:</i>			
Estimated coefficient of female dummy variable	-.235 (.001)	-.196 (.002)	-.180 (.002)
Relative contribution to wage gap	.628	.524	.481
Estimate relative to specification using percent female variables	.96	.93	.91
<i>With controls, full sample:</i>			
Estimated coefficient of female dummy variable	-.170 (.002)	-.142 (.002)	-.130 (.002)
Relative contribution to wage gap	.453	.380	.347
Estimate relative to specification using percent female variables	1.29	1.37	1.43
<i>With controls, restrict minimum size of job cell to 10 workers:</i>			
Estimated coefficient of female dummy variable	-.142 (.002)	-.112 (.003)	-.107 (.003)
Relative contribution to wage gap	.386	.330	.326
Estimate relative to specification using percent female variables	1.08	1.08	1.18

All specifications include the same controls as in column (6) of Table 4. The industry, occupation, and establishment effects are all absorbed in job-cell dummy variables, which are incorporated by forming within-job cell deviations.

Table 11: Estimated Log Wage Differentials by Sex, and Percent Female in Occupation, Industry, Establishment, and Job Cell, With Separate Wage Equations for Men and Women

	Men	Women	Men	Women
	(1)	(1')	(2)	(2')
Number of occupations	72	72	491	491
<i>Coefficient estimates:</i>				
% female in occupation	-.122 (.004)	.059 (.004)	-.129 (.004)	.021 (.003)
% female in industry	-.313 (.007)	-.116 (.007)	-.307 (.007)	-.090 (.007)
% female in establishment	-.127 (.006)	-.157 (.006)	-.130 (.005)	-.199 (.006)
% female in job cell	-.008 (.004)	-.220 (.004)	-.0004 (.004)	-.191 (.004)
R <sup>2</sup>	.379	.367	.379	.366
N	339,041	298,677	339,041	298,677
<i>Relative contribution to wage gap, male/female wage structure as no-discrimination wage structure</i>				
Female	.528/.336		.517/.292	
% female in occupation	.111/-.053		.136/-.023	
% female in industry	.208/.077		.203/.059	
% female in establishment	.115/.142		.118/.180	
% female in job cell	.014/.386		.001/.379	

Specifications include the same controls as in column (6) of Table 4. In calculating the decompositions, the coefficients on the number of children variables are treated as zero for men. When reporting the relative contributions, the entry before (after) the '/' uses the male (female) wage structure as the no-discrimination wage structure.

Table 12: Estimated Log Wage Differentials by Sex, and Percent Female in Occupation, Establishment, and Job Cell, Specific Industries

*A. Miscellaneous Plastics Products, Production and Maintenance Workers*

	<u>NWECD data</u>			<u>Groschen's results</u>
	(1)	(2)	(3)	(4)
Number of occupations	13/3	72/20	491/86	42
<i>Relative contribution to wage gap:</i>				
Female	.459	.376	.264	.067
% female in occupation	.190	.124	.223	.488
% female in establishment	.116	.139	.170	.121
% female in job cell	.236	.361	.342	.325
N		3,106		70,355
Number of establishments		324		876

*B. Nonelectrical Machinery, Production and Maintenance Workers*

Number of occupations	13/3	72/23	491/123	77
<i>Relative contribution to wage gap:</i>				
Female	.480	.380	.336	.010
% female in occupation	.102	.110	.088	.406
% female in establishment	.376	.308	.322	.480
% female in job cell	.042	.202	.254	.124
N		12,356		54,838
Number of establishments		734		795

*C. Banking, Nonsupervisory Workers*

Number of occupations	13/11	72/30	491/68	48
<i>Relative contribution to wage gap:</i>				
Female	.488	.531	.466	.024
% female in occupation	.309	.215	.289	.710
% female in establishment	.022	.029	.038	.048
% female in job cell	.182	.226	.208	.218
N		2,573		74,480
Number of establishments		440		579

All controls included in early tables are excluded here, to maintain comparability with Groschen's results; she does not have data on any of these controls. The results with the NWECD data were qualitatively similar including the human capital controls. Column (4) is from Groschen (1991),

Tables 1 and 2. In the rows labeled "numbers of occupations" we first report the number of occupations used in the occupational classification (corresponding to the earlier tables), and then the number of occupations represented in the data for the specific industry and occupations considered.

Table 13: Comparison of Sex Segregation in Job Cells

---

<i>A. Miscellaneous Plastics Products, Production and Maintenance Workers</i>		
	<u>NWECD data</u>	<u>Groschen's data</u>
	(1)	(2)
Number of occupations	86	42
Mean difference, women - men, in % female in job cell	.731	.761
<i>B. Nonelectrical Machinery, Production and Maintenance Workers</i>		
Number of occupations	123	77
Mean difference, women - men, in % female in job cell	.646	.646
<i>C. Banking, Nonsupervisory Workers</i>		
Number of occupations	68	48
Mean difference, women - men, in % female in job cell	.740	.505

---

See notes to Table 12.



Table 14: Estimated Log Wage Differentials by Sex, and Percent Female in Occupation, Establishment, and Job Cell, Specific Industries, Using Main IWS Occupation Categories

<i>A. Miscellaneous Plastics Products</i>			
	<u>NWECD data, IWS occupations</u>	<u>NWECD data, all production and maintenance workers</u>	<u>Groschen's results</u>
	(1)	(2)	(3)
Number of occupations	491/39	491/86	42
<i>Relative contribution to wage gap:</i>			
Female	.195	.264	.067
% female in occupation	.241	.223	.488
% female in establishment	.163	.170	.121
% female in job cell	.401	.342	.325
N	1,771	3,106	70,355
Number of establishments	227	324	876
<i>B. Nonelectrical Machinery</i>			
Number of occupations	491/26	491/123	77
<i>Relative contribution to wage gap:</i>			
Female	.377	.336	.010
% female in occupation	.076	.088	.406
% female in establishment	.200	.322	.480
% female in job cell	.347	.254	.124
N	5,626	12,356	54,838
Number of establishments	451	734	795
<i>C. Banking</i>			
	<u>NWECD data, IWS occupations</u>	<u>NWECD data, all nonsupervisory workers</u>	<u>Groschen's results</u>
	491/13	491/68	48
Number of occupations			
<i>Relative contribution to wage gap:</i>			
Female	.437	.466	.024
% female in occupation	.363	.289	.710
% female in establishment	.019	.038	.048
% female in job cell	.182	.208	.218
N	1,830	2,573	74,480
Number of establishments	390	440	579

---

See notes to Table 12. Estimates in columns (2) and (3) are reproduced from columns (3) and (4) of Table 12.

Table 15: Comparison of Coefficient Estimates Underlying Decompositions

<i>A. Miscellaneous Plastics Products, Production and Maintenance Workers</i>			
	NWECD data	Groshen's results	
	(1)	(2)	(3)
Number of occupations	20	86	42
Female	-.120 (.023)	-.084 (.027)	-.016 (.003)
% female in occupation	-.311 (.041)	-.438 (.044)	-.242 (.004)
% female in establishment	-.197 (.037)	-.241 (.035)	-.099 (.004)
% female in job cell	-.187 (.034)	-.150 (.035)	-.103 (.005)
<i>B. Nonelectrical Machinery, Production and Maintenance Workers</i>			
Number of occupations	23	123	77
Female	-.150 (.014)	-.133 (.016)	.003 (.004)
% female in occupation	-.362 (.026)	-.169 (.023)	-.452 (.006)
% female in establishment	-.357 (.026)	-.373 (.024)	-.330 (.007)
% female in job cell	-.143 (.026)	-.156 (.025)	-.058 (.007)
<i>C. Banking, Nonsupervisory Workers</i>			
Number of occupations	30	68	48
Female	-.299 (.034)	-.263 (.039)	-.009 (.003)
% female in occupation	-.588 (.052)	-.654 (.046)	-.685 (.008)
% female in establishment	-.102 (.060)	-.135 (.058)	-.385 (.012)
% female in job cell	-.198 (.048)	-.158 (.050)	-.160 (.008)

See notes to Table 12. The number of occupations listed is the number represented in the data for the specific industry and occupations considered. Specifications correspond to those in columns (2), (3), and (4) of Table 12, respectively.